DOCUMENT RESUME

ED 295 926 SP 030 297

AUTHOR Murray, Stephen L.
TITLE Synthesizing Teacher Testing Policy Options.
INSTITUTION Northwest Regional Educational Lab., Portland, OR. Assessment and Evaluation Program.
PUB DATE Oct 86
NOTE 38p.
PUB TYPE Reports - Descriptive (141)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS College Entrance Examinations; Elementary Secondary Education; *Policy Formation; *Teacher Certification; Teacher Characteristics; Teacher Effectiveness; *Teacher Evaluation; Teacher Promotion; *Testing

ABSTRACT

This paper examines key points to consider in developing policy on teacher testing. A summary is presented of the four underlying purposes that may be served, in part, by testing: (1) limiting the number of incompetent teachers; (2) encouraging teacher professionalism; (3) promoting public confidence in teachers; and (4) promoting excellence in education. The six types of decisions that provide the potential organizational context for testing are discussed: (1) admitting candidates to teacher preparation programs; (2) licensing teachers as sufficiently competent to teach; (3) certifying teachers; (4) selecting teachers for specific positions; (5) relicensing practicing teachers; and (6) granting promotions, rewards, or special status. The individual characteristics that may be considered in each decision class are briefly summarized. A discussion is offered on the sources of information about the individual characteristics that may be used in assessing teachers. Emphasis is placed on the different applications of teacher testing and incorporating newer policy options introduced in the Holmes Group Report and "A Nation Prepared" (Carnegie Forum). (JD)

THE Northwest Regional Educational Laboratory

PROGRAM REPORT

SYNTHESIZING TEACHER TESTING POLICY OPTIONS

By

Stephen L. Murray

For

Evaluation and Assessment Program

October 1986

Northwest Regional Educational Laboratory
101 S.W. Main Street, Suite 500
Portland, Oregon 97204

SYNTHESIZING TEACHER TESTING POLICY OPTIONS

Stephen L. Murray

Evaluation and Assessment Program
Northwest Regional Educational Laboratory

October 1986

SYNTHESIZING TEACHER TESTING POLICY OPTIONS

Stephen L. Murray

Northwest Regional Educational Laboratory

Introduction

Over the past several years, coinciding with a general movement to improve educational quality, the testing of prospective and experienced--already certified--teachers has increased significantly. A report by the Educational Testing Service (ETS) reveals that all but five states mandate, or have plans to mandate, the testing of prospective teachers (Anrig, 1986). As of the summer of 1986, almost one-half of the states were considering changes in their policies governing the use of tests to certify teachers.

Since 1979, the annual Gallup Poll of the public's attitudes toward public schools has shown strong support for teacher testing.

- In 1979, 85 percent of the general public supported using state controlled tests to certify prospective teachers for those subjects in which they planned to give instruction (Gallup G., 1979).
- In 1984, 89 percent of the general public favored such state controlled tests (Gallup G., 1984).
- In 1985, 89 percent of the general public favored requiring all teachers to pass a basic competency test to measure such things as their general knowledge and ability to think before they are hired by a school district (Gallup A., 1985).
- In 1986, 85 percent of the general public supported requiring experienced teachers to periodically pass a statewide basic competency test in their subject area or areas (Gallup A., 1986).

Although the questions have varied somewhat from year to year, public attitude clearly supports teacher testing as a way to control the quality of the teaching force. Further, as evidenced by their responses to the 1986 poll, the public would not exempt experienced teachers from such testing.

In the late spring of 1986, two reports fueled the public's interest in improving the quality of the teaching force. The first report, Tomorrow's Teachers: A Report of The Holmes Group (1986), was developed by a consortium of education deans. The three-tier system of teacher licensing and certification proposed in the Holmes Report includes:

1. Instructors
2. Professional Teachers
3. Career Professionals

Assessment, including testing, would be an important method of certifying teachers at each of these three tiers.

The second report, "A Nation Prepared: Teachers in the 21st Century," prepared by the Carnegie Forum on Education and the Economy's Task Force on Teaching as a Profession also proposed restructuring the teaching force (1986). Their proposal suggests four teaching levels including:

1. Licensed Teachers
2. Certified Teachers
3. Teachers with Advanced Certificates
4. Lead Teachers

The Carnegie Task Force report recommends creating a National Board for Professional Teaching Standards to oversee a rigorous national teacher certification system. The National Board would set high standards supported by a three-stage assessment strategy.

Certification through the National Board, as proposed in the Carnegie Plan, would be voluntary and independent of state licensing. The Carnegie Plan, however, sees the National Board certification system as going beyond state licensing requirements and suggests that states may wish to waive their specific licensing requirements for teachers who have been Board certified. The Carnegie Foundation has funded efforts to establish such a Board and begin developing the assessment tools necessary to support the professionally oriented certification testing it advocates.

In an earlier paper, we urged that teacher testing policies be rooted in clear and consistent underlying purposes (Murray, 1986). We suggested a framework of four purposes which underlie current practice and recent teacher testing proposals. These purposes are:

1. Limiting the number of incompetent teachers
2. Encouraging teacher professionalism
3. Promoting public confidence in teachers
4. Promoting excellence in education

Given an assessment of these broad purposes, a second consideration in developing policy options is to examine the implications of introducing tests, altering their content or altering the way the tests are used to contribute to each of the following six types of institutional decisions:

1. Admitting candidates to teacher preparation programs
2. Licensing teachers as sufficiently competent to teach
3. Certifying teachers
4. Selecting teachers for specific positions
5. Relicensing practicing teachers
6. Granting promotion's, rewards, or special status

We have added a sixth decision type, certifying teachers, to reflect the distinction between licensing and certification emphasized by the proposals in the Holmes Report and the Carnegie Plan. Given a clear and consistent set of purposes and types of decisions to be supported with test data, a reasonable third step will be to review the requirements of tests in relation to the intended purposes and interpretations implied. The primary concerns for tests will be validity, reliability, lack of bias and incremental utility over other methods for making decisions.

A synthesis of best practice should be based on a careful analysis of the consequences of adopting policy actions. But, because many of the policy options are unprecedented, there is a scant empirical base for such an analysis. Nonetheless, as new proposals emerge, the need for information to guide new policy initiatives is growing. This paper and an earlier paper (Murray, 1986) address key points to consider in developing policy on teacher testing. The next two sections summarize the four underlying purposes that may be served, in part, by testing and the six types of decisions that provide the potential organizational context for teacher testing. The third section summarizes the individual characteristics that may be considered in each decision class, and the fourth section briefly discusses the sources of information about the

individual characteristics that may be used in assessing teachers. The earlier paper (Murray, 1986) discusses issues of test validity and fairness (or lack of bias). The present paper places more emphasis on the different applications of teacher testing and incorporating newer policy options introduced by the Holmes Group and the Carnegie Plan.

Purposes Underlying Teacher Testing

First, testing is a means to limit the number of incompetent teachers. Vorwerk and Gorth (1986) state that the primary outcome of every state's teacher certification system is to protect the public from teachers who lack basic communication skills and essential subject matter knowledge in those areas they will be called upon to teach. Testing is an objective means to determine whether a candidate possesses the skills and knowledge needed to minimally fulfill job requirements. Unless a teacher candidate demonstrates minimum competence, he or she will be prevented from entering the classroom as a licensed teacher.

A second underlying purpose for teacher testing is to encourage teacher professionalism. Shanker (1985) and Schulman (1986) have argued that to support teacher professionalism, assessment and testing systems should be modeled after such certification systems as nongovernmental medical specialty boards. Control of the testing and certification processes would be in the hands of the community of professional teachers rather than the state. The content of tests for teachers would reflect the expert knowledge required to match the complex job requirements of teaching and contribute to a greater public valuing of teaching as a

profession (Schulman, 1986). Certification would become a way to recognize competence beyond the minimum necessary to function on the job.

The distinction between limiting incompetent teachers and encouraging professionalism, raises what can be a source of confusion in the literature on teacher testing policy. Licensing, a more restrictive function than certification, is formally defined "as a process by which an agency of government grants permission (emphasis added) to an individual to engage in a given occupation upon finding that the applicant has attained the minimal degree of competency required to ensure that the public health, safety, and welfare will be reasonably well protected " (U.S. Department of Health, Education, & Welfare, 1977, p. 4).

Certification, on the other hand, "is the process by which a governmental or nongovernmental agency grants recognition (emphasis added) to an individual who has met certain predetermined qualifications set by a credentialing agency. Unlike licensing, a certification law does not prohibit uncertified individuals from practicing their occupations" (Shimberg, 1981, p. 1138).

Where state laws prohibit one who is not "certified" from teaching in that state, the term teacher certification stands for the more restrictive licensing function. Both the Holmes Report and the Carnegie Plan call for entirely new certification systems that are independent of state efforts to protect the public welfare and which would give recognition to higher levels of competence. Certification requirements transcend licensing requirements. To reflect this distinction in the present paper, we use the term licensing to refer to the state function of determining whether a teacher will be granted permission to teach in that state. Most states refer to this practice as certification.

A third reason for implementing teacher testing policies is to promote public confidence in the teachers as a group. Gallup polls have shown strong public support for testing prospective and experienced teachers in the subject areas they expect to teach or do teach and for other skills and knowledge judged to be important for functioning effectively as a teacher (Gallup G., 1979, 1984; Gallup A., 1985, 1986). The percentage favoring state controlled teacher testing ranges from 85 percent to 89 percent. The last two polls gave similar support to testing experienced teachers and to testing teachers as a condition of employment. This is not to say, however, that testing teachers will, in fact, increase public confidence in teachers.

A fourth reason for teacher testing is to promote excellence in education. Teacher testing programs that are part of more general educational reforms promote excellence indirectly by symbolizing higher standards of performance. Excellence prevails when the best teachers are hired, when superior teachers are recognized and when good teachers are encouraged to stay with the profession.

There are two major limitations of testing programs designed to support such higher level purposes. First, not all teacher testing policies will serve all purposes equally well. More importantly, some policies may lead conflict between fundamental policy purposes. For example, using state controlled tests to screen out experienced teachers who lack minimally necessary communication skills and subject matter knowledge for a beginning teacher may detract from, rather than promote

teacher professionalism. Such a policy, if deemed necessary, may also raise public doubt about the quality of the teaching force.

Second, given the underlying purposes that teacher testing is expected to support, it is important to keep in mind that testing is only one tool to improve the quality of the teaching force and to improve education in general. Moreover, to be effective, it must be accompanied by other policy initiatives. For example, testing teachers is only one aspect of a policy to limit incompetence. One reason is that tests of subject-matter knowledge and basic communications skills are not appropriate for directly assessing other areas in which teacher competence may be a concern. Bridges (1986), for instance, reports that the leading cause for teacher dismissal--in over seventy years of research--is weakness in maintaining student discipline. Problems in maintaining rapport with other teachers and parents, and failure to produce intended classroom outcomes are other frequent causes for teacher dismissal. Clearly, tests of knowledge are not designed to predict the ability to maintain discipline and rapport, or to produce intended classroom outcomes.

Objective paper and pencil tests are only one form of assessment designed to identify only one form of competence--lack of knowledge. Forms of assessment and evaluation other than paper and pencil testing and appropriate validation are needed to measure these other areas of competence. More direct observations are needed to validly assess these characteristics.

Policy must be clear on the purposes to be achieved and the appropriate timing for assessing different content domains. Testing teachers for relicensing, for example, is not likely to detect a large proportion of teachers who warrant more careful evaluation. For example, nearly 97 percent of the educators who took the Texas Examination of Current Aministrators and Teachers (TECAT) passed it on the first administration (Rodman, 1986), and over 99 percent passed by the second administration. We speculate from Bridge's research that the TECAT identified only a small percentage of the incompetent teachers as administrators. For this reason, the general public may be skeptical of test results that identify such small percentages of teachers as incompetent. Testing sampled only a narrow aspect of teacher competence. In other words, as a screen for limiting incompetence, testing teachers was too limited.

The Washington Post raised just such a concern in its editorial, "Grading teachers: Tough Enough?" (Washington Post, 1986). The Washington Post drew attention to the result that while 45 percent of first year teachers in Virginia failed the first round of evaluation with a new system of evaluating classroom performance, 98 percent passed in two attempts. The Post commented:

> A good standard will weed out those who should not be in the state's classrooms, while improving the performance of those who had trouble on the evaluations, but can--with help--still become good teachers. But evaluations will serve no purpose at all if they are easy enough for everyone to pass (p. B6).

Teacher testing influences the attainment of the higher level purposes through its use in the institutional decision making. In the next section, we describe the common uses of tests in the context of institutional use for decisions--which is what policy generally affects most directly.

Decisions Supported by Teacher Testing

Institutional Decisions

Testing teachers can support institutional or individual decisions (Cronbach and Gleser, 1965). Most teacher testing policies, however, support institutional decisions or recurrent choices made about individuals by an agent or agents acting on behalf of an institution. The individuals about whom choices are made include applicants for teacher training, those who are trained to be teachers and are seeking licensing, others who are seeking certification or jobs, and currently certified teachers. State agencies, colleges or universities involved in teacher training, and local school districts are the primary agencies involved in using test results to support decisions about these individuals. The Carnegie Plan would create a new nongovernmental body, representing the teaching profession, to operate a professional certification system.

Murray (1986) proposed a framework of five institutional decisions that currently utilize teacher testing. Analysis of the Carnegie Plan suggests adding professional certification as a sixth type of institutional decision which expands the list to include:

1. Admitting candidates to teacher preparation programs
2. Licensing teachers as sufficiently competent to teach
3. Certifying teachers
4. Selecting teachers for specific positions
5. Relicensing practicing teachers
6. Granting promotions, rewards or special status

In addition to these uses of data for decisions about individuals, licensing test results for teachers are often used to evaluate teacher training programs. Each of these decision types, as well as the underlying policy rationale for teacher testing initiatives, places specific demands on testing and evaluation design.

Admission to Teacher Preparation Programs. The admissions decision, which is typically made after the candidate has completed two years of college (Schalock, 1979), is intended to select those who will successfully complete the preparation program and who will subsequently become licensed and certified to teach. In other words, to select those who show promise as teachers. Usually, the number of teacher candidates admitted to an institution's preparation program depends upon the number of persons the training program can accommodate. The admissions decision, therefore, operates with a selection quota and is norm referenced.

Generally, the weight given to test-derived information is not specified and may be difficult to determine in practice. A policy issue of some importance is, Under what conditions will nontest data be allowed to compensate for poor test performance? Can nontest data outweigh even the poorest test performance, or is there an absolute cutoff score on the

test? Use of an absolute cutoff score means that only those above a predetermined score are given further consideration for admission. A very low absolute cutoff, one that a high percentage of applicants can be expected to pass, may give the test less weight than a decision rule allowing poor test performance to be compensated for with other information. In other words, using an absolute cutoff score, by itself, does not determine the importance (weight) of the test in making the decision.

When admissions decisions are based on a quota, the goal is to select the best of many applicants. The admissions test must reliably discriminate between applicants across a broader range of talent than will licensing tests, which need only to discriminate between those who possess the minimum required level of knowledge and those who do not. On the other hand, when an admissions test is used only as a preliminary hurdle, the test needs only to identify candidates with minimally acceptable performance.

Currently, teacher training applicants in most states are not expected to have pedagogical knowledge or a high level of subject matter expertise, because they are tested two years after beginning college and prior to teacher training. The Holmes Report and the Carnegie Plan would postpone teacher training until teacher training applicants complete their undergraduate training, making it realistic to require tests of subject matter knowledge as part of the information for admission.

Under current conditions, tests used to support admission decisions typically measure basic literacy or academic skills and include such tests as:

1. The Pre-Professional Skills Test (PPST)
2. The California Basic Educational Skills Test (CBEST)
3. The Alabama English Language Proficiency Test (ELP)
4. The Connecticut Competency Examination for Prospective Teachers (CONNCEPT)
5. The California Achievement Test (CAT)
6. The Scholastic Aptitude Test (SAT)

The PPST, CBEST, CONNCEPT, and ELP are designed specifically for teaching candidates, the CAT measures academic achievement, and the SAT measures academic aptitude and is used for general college admissions decisions. The PPST, which measures basic proficiency in reading, writing and mathematics, was developed by the Educational Testing Service (ETS) to test content similar to that of the National Teacher's Examination (NTE) but which would be appropriate for teacher training applicants not yet exposed to specific teacher training. The CBEST is similar to the PPST, having been developed to the specifications of the California State Department of Education by the ETS. The major difference between the PPST and the CBEST is that the CBEST includes two written essays while the PPST has only one essay and an objective multiple choice test of writing knowledge.

The ELP and the CONNCEPT are custom-designed, criterion-referenced tests. The ELP measures competencies needed for successful completion of course work in the teacher education program and for effective classroom teaching (Baker and Fennel, 1986). The CONNCEPT was designed with the same general goal in mind (Pecheone, Tomala, and Forgione, 1986).

Initial Licensing. The decision to license one to teach follows teacher preparation (which may include practice teaching or an internship) and is the responsibility of a state governmental agency such as a state department of education or a licensing agency. The typical goal of licensing is to validly, fairly and efficiently identify candidates minimally competent to teach in the state. The licensing decision is criterion referenced in that there is not a fixed quota of positions to fill. A licensing test determines whether candidates are qualified to teach in the state, not whether they will be more or less successful.

Theoretically, the percentage of candidates passing the test could range from 0 percent to 100 percent as long as the test validly and fairly discriminates between those who possess the minimally required knowledge and those who do not. The key, therefore, is to determine an objective standard of what is minimally acceptable. Problems arise when the demand for licensed teachers exceeds the supply. When this happens, emergency certificates may be granted. Both the Holmes Report and the Carnegie Plan, however, caution against lowering standards for licensing when demand exceeds supply. To keep the supply of teachers equal to or greater than the demand, they each argue that it is essential to increase incentives to attract more people to enter the teaching force. This policy position clearly illustrates the idea that for teacher testing policies to be effective, they must exist within a broader network of initiatives.

The logic of licensing tests requires that specific cutoff scores be set for each test. Candidates scoring above the cutoff are licensed to teach while those scoring below the cutoff must either retake and pass

the test, or fail to be licensed. The test, therefore, should be effective at discriminating between acceptable and unacceptable candidates. Because of this requirement and the fact that any test is only a limited sample of behavior, tests for licensing may not do a particularly good job of discriminating across a wide range of knowledge. One implication for policy is that a test that is both valid and efficient for licensing beginning teachers may not be a good test for identifying teachers with superior knowledge in the area tested. Whether it is or is not depends on the content sampled by the test and the validation methods followed.

Licensing tests, generally paper and pencil tests, should measure content knowledge that the state has demonstrated is essential for beginning teachers in that state. As such, licensing is strictly a state level institutional decision, the purpose of which is to limit the number of incompetent teachers. Each state has its own licensing requirements, although many states have reciprocal agreements to deal with teachers prepared out of the state in which they apply for a license. Most recent teacher testing policy developments have been in the area of teacher licensing.

One way in which testing policies licensure differ is in how they provide for candidates to retake the test and what assistance, if any, is provided to help candidates pass the test. Policies differ in the amount of time allowed between testings and the number of times a candidate can retake the test. A second important difference in the testing policies of different states is in the level of difficulty reflected in the cutoff score. While stringent cutoff scores imply more rigorous standards for who will be licensed to teach, states with more rigorous standards may be

those who have a larger supply of teacher candidates, or who more easily grant provisional licenses.

Information used to support licensing decisions, which is not limited to test information, may include:

1. General knowledge
2. Knowledge of teaching methods
3. Knowledge in subject(s) the candidate plans to teach
4. Communication skills
5. Successful completion of an approved teacher training program
6. Commitment to teaching
7. Acceptable trial performance of teaching functions

Assuming that the candidate has completed training, it is reasonable to expect information more specific to functioning as a teacher and not to require information about basic knowledge or skills. Basic academic skills measures used to screen students for admission to teacher training programs have already been used incidentally in licensing. But, of course, one of the reasons for testing basic knowledge and communication skills is the concern that teacher candidates from institutions outside those under the control of the state may not have been subject to comparable quality screens. One of the specific reasons for standardizing teacher licensing testing, therefore, has been to control for a lack of standard information on the quality of teacher candidates.

Tests used to license teachers include:

1. The National Teacher's Examination (NTE)
2. The Pre-Professional Skills Test (PPST)

3. The California Basic Educational Skills Test (CBEST)

4. The Georgia Teacher Certification Tests

In addition to using these tests, a number of other states have developed their own licensing tests, contracting to such agencies as National Evaluation Systems (NES) and the Instructional Objectives Exchange (IOX). Oklahoma's program includes criterion-referenced tests for more than 75 different subject areas (Folks, 1986).

Of the tests used for licensing, the NTE, which is published and managed by ETS, is used most widely. The NTE testing program, which began in 1940, comprises objective, standardized measures of academic preparation for teaching. The primary purpose of the NTE battery was to allow school systems "to evaluate the achievement of individuals from different colleges and universities which may have dissimilar standards and grading practices." (Rosenfeld, Thornton, & Skurnik, 1986, p. I-1). Recently revised, the NTE Core Battery includes tests of professional knowledge, general knowledge and communications skills. The NTE Specialty Area Tests measure 27 content areas (two additional Specialty Area tests are available for limited use).

As of the summer of 1986, 17 states used the NTE Core Battery for teacher licensing. Thirteen of these states also used the NTE Specialty Area Tests for licensing. To support the legal use of the NTE in a state, the state must validate the NTE content and set appropriate cutoff scores. Validity studies must establish the content validity of the NTE and specific cutoff scores in relation to the minimum knowledge required to function as a beginning teacher in that state.

Modified tests and custom-made tests are used in states where

resources were available for their development and where policy objectives called for a state specific testing program. Whether a state decides to use a published test or develop its own test, however, they are responsible for validating the test for use in their state and setting state standards of minimally acceptable performance. Unlike tests used for admissions to teacher training programs, tests used for licensing must be validated in terms of job relevance, and cutoff scores must be based on what is minimally required to perform as a beginning teacher in a state.

Professional Certification. This third category of use reflects the need to clearly distinguish between state controlled licensing and professionally controlled certification. The three-tier system of teacher licensing and certification proposed in the Holmes Report includes:

1. Instructors
2. Professional Teachers
3. Career Professionals

Before they are certified, all Instructors would be required to pass a written test in each subject area they will teach. Further,

> this exam should test for their understanding of the structure of the discipline, and the tenets of a broad liberal education. They should additionally pass a general test of their reading and writing ability, and a test of the rudiments of pedagogy. These tests would assess reasoning as well as specialized knowledge, general information, and memory. They should be sufficiently difficult so that many college graduates could not pass: (p. 11).

Professional Teachers would have to pass the same subject area tests and general reading and writing tests as Instructors. In addition, they

would have to pass examinations in pedagogy and human learning. Although the Holmes Report proposes additional requirements for Career Teachers, it does not propose additional testing.

The Carnegie Plan calls for a three stage assessment strategy. The first stage, focusing on subject matter content, would coincide with graduation from college. The second stage, centering on student mastery of subject matter covered in professional education coursework, would be applied whenever the student is ready. The third stage, based on extensive observation of the candidates actual teaching, would complete the assessment for certification.

Selection for Teaching Positions. A fourth use of teacher testing is in the process of selecting applicants for a teaching position. The norm is for local school policies, rather than the state, to prescribe how teachers will be selected for local positions. One state, Hawaii, requires use of both the NTE Core Battery and the Specialty Area tests as part of the information considered in hiring teachers. Selection decisions are norm referenced as they focus on selecting the best of an available applicant pool.

Use of a test to support hiring decisions is also subject to the Uniform Guidelines for Employee Selection Procedures developed by the Equal Employment Opportunity Commission (EEOC) (EEOC et al., 1978). The Uniform Guidelines for Employee Selection include the following provisions:

1. Empirical data should be made available to establish the predictive validity of a test, that is, the correlation of test performance with job-relevant work behaviors; such data should be collected according to generally accepted procedures for establishing criterion-related validity.

2. Where predictive validity is not feasible, evidence of content validity (in the case of job knowledge proficiency tests) may suffice as long as appropriate information relating test content to job content is supplied.

3. Where validity cannot otherwise be established, evidence of a test's validity can be claimed on the basis of validation in other organizations as long as the jobs are shown to be comparable and there are no major differences on context or sample composition.

4. Differential failure rates (with consequent adverse effects on hiring) for members of groups protected by Title VII constitute discrimination unless the test has been proven valid (as defined above) and alternative procedures for selection are not available.

5. Differential failure rates must have a job-relevant basis and, where possible, data on such rates must be reported separately for minority and nonminority groups.

Relicensing Practicing Teachers. A fifth decision for which test data may be considered is to relicense currently licensed teachers. Relicensing is also a criterion-referenced decision. Three states, Arkansas, Georgia and Texas, have implemented programs to test teachers who are already licensed. The Texas program, which uses the Texas Examination of Current Administrators and Teachers (TECAT), also includes testing administrators. These programs are subject to considerable controversy which stems, in part, from the confusion between the underlying purpose of a licensing program, which is to limit the number of incompetent teachers and the more broadly conceived programs to promote educational excellence. States which require testing for relicensing are forced to acknowledge that either job requirements have changed or that previous licensing standards were unacceptably low. Notably, ETS has forbidden use of the NTE for relicensing decisions. Popham (1985) argues persuasively that testing in basic skills of reading, writing, or mathematics is a viable approach to identifying teachers who are deficient in these areas. What may decide the issue is

the percentage of teachers who would be identified through such testing and the costs of testing versus other alternatives to identifying problem teachers.

Career Advancement. A sixth decision for which test data may be considered is to support career ladder programs in which teachers are given opportunities to increase their level of professional responsibility by taking on special assignments. Florida and Tennessee are two states using tests within career ladder programs. (See Fisher, Fry, Loewe and Wilson (1985) for a description of the Florida program.) As with using tests for relicensing, there is little experience with using tests for career ladder programs. Controversy over these programs has caused delays in implementation and is reported to have negatively affected teacher job satisfaction. The problems are more deeply seeded than those that may be related to testing and relate to the more general concerns with the fair and valid administration of merit pay (Murnane and Cohen, 1986) and assigning differentiated responsibilities. In practice, using tests for career advancement has elements of both norm-referenced decisions and criterion-referenced decisions. However, assuming that there are a limited number of positions with enhanced responsibilities, career advancement decisions are fundamentally norm referenced.

Individual Decisions

Individual decisions relate to setting personal goals and direction rather than making routine institutional decisions such as admissions, licensing and selection. A teaching candidate may decide, for example, to concentrate on teaching in a specific content area to take advantage

of an assessed strength, or the candidate may decide to remedy a weakness revealed by a test. Shimberg (1981), discusses the use of self-assessment testing as a tool to ensure the continued competence of practicing professionals. In self-assessment testing, the professional voluntarily takes tests with the assurance that no one else will know the results. The underlying concept is that practitioners may be unaware of their own areas of weakness and that much of the anxiety and opposition that results from submitting to testing is eliminated or reduced. Individuals use their results to plan their own refresher training. Group data, although not representative of the population, may be used to plan educational programs for the professional group. Self-assessment testing can be combined with relicensing testing to give practitioners a way to assess weaknesses prior to the "official testing." Self-assessment has received little attention, but does have potential for improving teacher quality.

Evaluation of Teacher Training Programs

While evaluating teacher training programs does not fit into the framework of institutional or individual decisions, we mention this use briefly because teacher testing results used for licensing are sometimes used to evaluate the quality of teacher training programs.

If the policy objective is to evaluate and improve the quality of teacher training programs rather than focus on individual decisions, then there is a need for evaluation strategies that appraise the quality of the program independent of their selection practices. Teacher training programs are likely to differ in their capacity to attract capable

students. Those institutions whose students have a higher success rate on certification examinations may simply be those who are more selective.

On the other hand, institutions that take greater risks (by choice or by chance) may have lower passing rates because of their recruitment and selection rather than their ability to train. Given forecasts of shrinking proportions of minority candidates who will qualify as teachers, some consideration should be given to offering incentives to schools that will take greater risks and who are effective. The quality of the instructional program offerings should, however, be fairly evaluated. That is, one should press to determine whether their programs are effective by using more information than passing scores on a test.

An issue that has received more attention is that some content tested is not taught in teacher education courses. The concern has been that including such content would result in an unfair evaluation of the teacher training program.

To summarize, teacher testing programs should clearly specify the institutional and individual decisions they will serve and they should be justified in terms of more basic purposes such as those discussed in the previous section.

A Summary of Characteristics Measured

Table 1 summarizes individual characteristics considered in each of the six institutional decision categories discussed in the previous section. Individual characteristics in Table 1 include those used currently and those suggested in the Holmes Report and the Carnegie Plan. The two columns of Table 1 list those characteristics which may be

measured with objective, paper and pencil tests and those which may be most effectively assessed through other means (e.g. records, observation, interviews).

Recall that the primary admissions objective is to select those candidates who will successfully complete teacher training. Success is most often defined in such terms as completing the teacher training program, qualifying for a teaching license, qualifying for teaching positions for which the candidate has been prepared and entering teaching as a career. Longer term criteria for success, such as effectiveness as a teacher, are rarely considered as criteria for admissions decisions or for validating tests used for admissions (Schalock, 1979).

TABLE 1

Sample Types of Information for Decision Making

Decision	Test Content	Other Information
Admissions	Academic Aptitude Basic Literacy Reading Writing Mathematics Subject Area Knowledge	High School GPA Undergraduate GPA Interest in Teaching Teaching-like Experience Oral Communication Skills Speaking Listening
Licensing	General Knowledge Knowledge of Pedagogy Subject Area Knowledge Communication	Successful Teacher Training Commitment to Teaching Successful Trial Performance Oral Communication Skills Speaking Listening
Certification	Subject Area Knowledge Knowledge of Pedagogy Reading Proficiency Writing Proficiency Principles of Learning	Successful Trial Performance Successful Course Completion
Selection	Basic Literacy Reading Writing Mathematics Subject Area Knowledge General Knowledge Knowledge of Pedagogy Communication Skills	Successful Teacher Training Commitment to Teaching Oral Communication Skills Trial Performance Previous Experience
Relicensing	Basic Literacy Subject Area Knowledge	Satisfactory Performance Additional course work
Career Advancement	Subject Area Knowledge	Accumulated Experience Peer Evaluation Student Performance Classroom Observation

The need for information on academic aptitude recognizes that preparation requires success in the academic courses in the teacher training program. Information on candidate basic literacy identifies deficiencies in skill areas that the training program considers prerequisites and, therefore, does not address in its course of study. Subject area knowledge may be used when the teacher preparation program requires the candidate to have completed an undergraduate program before applying for teacher training. Information on candidate interest in teaching and prior experience in teaching-like activities (e.g. tutoring experience, volunteer work with children) may be appropriate to consider in admissions decisions. Whether changes in policies for testing teacher candidates are effective depends on whether newer practices will improve the rate of correct decisions.

The success of an admissions process is limited by the success of the recruitment process and the extent to which the process leads to admitting an acceptable number and proportion of minority candidates. The more successful teacher training institution will do better at attracting larger numbers of better qualified candidates. Any teacher training institution, however, has limited capacity to attract students. Prospective teachers must feel that entering the teaching force will be rewarding; that is, that there will be jobs available, pay will be acceptable and there will be opportunities for advancement.

The objective for licensing testing is to identify those candidates who possess the minimum knowledge and skills to perform successfully as a beginning teacher and to exclude all others. The judgment as to what knowledge and skills are needed must be based on evidence from an analysis of the general job requirements for a beginning teacher in a state. That job analysis, which frequently includes teacher surveys but

may also draw from other sources, supports claims for test validity. Table 1 includes a listing of individual characteristics frequently considered in licensing teacher candidates.

In theory, the success of a licensing program may be defined in terms of the percent of correct decisions it supports; that is, of all those licensing decisions made, what percent are correct? There are two types of correct decisions. The first correct decision is to license a candidate who does not constitute a threat to the public's welfare as a teacher and the second is to withold a license from one who does represent a public threat. Two incorrect decisions would be to license one who is a threat to the public welfare and to withold a license from one who poses no such threat. These theoretical outcomes bring out an important point. Because tests can only identify limited areas of threat to the public welfare (i.e. lack of knowledge), they are a necessary but insufficient basis for licensing. The policy implication for a state examining its licensing requirements would be to seek to determine whether lack of subject area competence represents a major threat to the public welfare or whether some area other than subject area knowledge is the primary threat.

Certification's basic purpose is to grant recognition to teachers qualified for responsibilities beyond those of the minimally qualified beginning teacher. The general content domains for certification tests have been suggested in the Holmes Report and the Carnegie Plan and are tied to differential levels of teaching responsibility. Such certification testing would be based on an idealized conception of the teaching force rather than empirical analyses of the factors distinguishing between teachers who actually have such differential responsibilities. Because the changes called for require reform in the

structure of teaching, validation programs for the new certification tests will need to extend into the period in which the different levels of teachers are recognized and incorporated into the design of teaching positions. Before such empirical analysis is possible, validation will necessarily be based on a rational analysis of the requirements underlying the different teaching levels proposed.

The success of certification testing programs will be intertwined with the success of attempts to reform teaching. For certification testing to be valued, the concepts that underlie it must be recognized in the design of teaching positions and in granting special status to teachers who have achieved higher levels of certification. Certification testing must be validated against other performance criteria of teaching effectiveness if it is to be accepted as moving teaching toward a profession.

Selection seeks to identify the best candidates for a limited number of teaching positions. As previously discussed, selection must conform to the Uniform Guidelines for Employee Selection (EEOC, 1978). Therefore, it is essential that any selection test be validated against job-relevant work behaviors. In the case of job-knowledge tests, evidence of content validity may suffice as long as test content can be related to job content. Job content will depend on the duties that a new teacher may be called upon to perform and will, therefore, depend on the needs of the hiring district. Table 1 suggests the individual characteristics that may be derived from such assessments of job relevance. For tests to contribute to the success of selection decisions, it is necessary that the rate of correct decisions is greater with the use of tests. If tests are used in prior screening (e.g. for

admissions, licensing and certification), it may very well be that additional testing may not contribute much to the rate of correct decisions in a school system.

Relicensing is meant to insure that teachers who have been licensed continue to maintain a level of knowledge and skill to perform at least at a level necessary to protect the pubic against incompetence. Generally, it is important to insure that teachers keep abreast of developments in the subject areas they are responsible to teach and that they perform required teaching duties in at least a satisfactory manner. Testing is one method of gathering information about individuals being considered for relicensing. However, testing seems far less important than other forms of evaluation more closely tied to performance.

Finally, decisions about who should receive career advances are intended to advance teachers who are more deserving based on their knowledge and performance. Career ladder programs attempt to create more opportunities for teachers to gain recognition and financial rewards for assuming greater responsibilities. Testing has been used as one tool to determine who qualifies for advancement. Table 1 lists some of the factors considered in making these types of decisions. Career ladder programs in Florida and Tennessee have been subject to implementation problems for various reasons only incidentally related to the use of tests for making the decisions. At least part of the difficulty in implementing these programs is that they call for restructuring teaching positions into a more hierarchical pattern. Career ladders and the professional certification movement share this concept of differential teaching roles that imply differences in status and pay. The major

difference is that the former have usually been controlled by the state and the latter are to be controlled by a National Board representing professional teachers.

Sources of Information About What to Measure

In terms of professional testing standards and legal requirements, job analysis is an essential basis for validating tests used to license, certify, select and promote teachers. Job analysis is intended to determine the minimum knowledge necessary to perform as a beginning teacher. It can also be used to determine the factors that distinguish a superior teacher from an adequate teacher. Job analysis used to validate teacher tests is usually limited to surveys of practicing teachers. It can, however, consist of a much greater variety of techniques.

Table 2 presents the four aspects of job analysis outlined in McCormick (1976). The first aspect lists types of job analysis information that may be collected. Note that knowledge applied and job-related knowledge and skills are two subclasses of information that might be the focus of a job analysis. Many other types of information such as job context and worker oriented activities (i.e teaching behavior) could be developed through a job analysis. In addition, eleven different methods of collecting job information are listed in Table 2. Limiting the methods of job analysis begs the question, What are the essential skills and knowledge of a beginning teacher? This is not to say that job analysis has not been applied to teaching. A great deal of descriptive research on teaching and teaching effectiveness qualifies as job analyses of teaching. Consequently, research on teaching could justifiably be included in a job analysis.

While teacher surveys of the importance of subject matter content are appropriate, they can, if used to the exclusion of other information, be misleading because they do not account for nonsubject matter determinants (e.g. teaching methods, materials, work context) of teaching. A recent survey of the National Science Teachers Association (NSTA) resulted in a significant reversal of their recommended coursework requirements for science teachers (Rothman, 1986). NSTA had advocated the need for greater training in specific disciplines as a basis for certification. They found, however, that only a small percentage of secondary-school science teachers had single discipline teaching assignments. Most science teachers teach several science subjects rather than specialize in a single discipline. NSTA has since concluded that science teacher training should be broadened rather than require a degree in a specific discipline.

The job of teaching is influenced by factors in addition to specific subject areas--school size, course requirements, course demand, etc. The issue is, How general should the certification demands be? Given limited resources, what options are there for efficient use of tests? One option could be to narrow what is tested by the licensing agency and attempt to improve the quality of testing as part of preparation. If the Carnegie Plan is realized, more sophisticated tests may be available, but they do not yet exist.

TABLE 2

Four Aspects of Job Analysis

Type of Job Analysis Information

- Work Activities
  - Job-oriented Activities
  - Worker-oriented Activities
- Machines, tools, equipment and work aides used
- Job Related Tangibles and Intangibles
  - Materials Used
  - Products
- Knowledge Applied
  - Services Rendered
- Work Performance
  - Work Measurement
  - Work Standards
  - Error Analysis
  - Other Aspects
- Job Context
- Personnel Requirements
  - Job-related Knowledge/Skills
  - Personal Attributes (Aptitudes, Interests etc.)

Form of Job Analysis Information

- Qualitative
- Quantitative

Methods of Collecting Job Information

- Observation
- Individual Interview with practitioners
- Group Interviews
- Technical Conference
- Structured Questionnaire
- Unstructured Questionnaire
- Diary
- Critical Incidents
- Equipment Design Information
- Recordings of Job Activities (e.g. video tape)
- Records

Agent used to Collect Job Information

- Individuals
  - Job Analyst
  - Supervisor
  - Incumbent
- Devices
  - Video tape

References

Anrig, G. R. (1986). Teacher education and teacher testing: The rush to mandate. Phi Delta Kappan, 67, 447-451.

Baker, C. C., & Fennell, Barbara (1986). The Alabama english language proficiency test: A criterion for admission to teacher education programs. In W. P. Gorth & M. L. Chernoff (Ed.), Testing for teacher certification (pp. 253-265).

Bridges, E. M. (1986). The incompetent teacher. Philadelphia: The Falmer Press.

Carnegie Forum on Education and the Economy's Task Force on Teaching as a Profession. (1986). A nation prepared: Teachers for the 21st Century. New York: the Carnegie Forum on Education and the Economy.

Cronbach L. J. & Gleser, G. C. (1965). Psychological tests and personnel decisions (2nd ed.). Urbana, IL: University of Illinios Press.

Equal Employment Opportunity Commission. Civil service Commission. Department of Labor. & Department of Justice. (1978, August) Adoption by four agencies of Uniform Guidelines on Employee Selection Procedures. Federal Register, 43,(166), 38290-38315.

Fisher, Thomas H., Fry, Betty V., Loewe, Kenneth L., & Wilson, Garfield W. (1985, Fall). Testing teachers for merit pay purposes in Florida. Educational Measurement: Issues and Practice, 4(3), 10-12.

Folks, John (1986). Stakeholders in teacher certification testing. In W. P. Gorth & M. L. Chernoff (Ed.), Testing for teacher certification (pp. 47-57).

Gallup, G. H. (1979). The 11th annual Gallup poll of the public's attitudes toward the public schools. Phi Delta Kappan, 61(1), 33-45.

Gallup, G. H. (1984). The 16th annual Gallup poll of the public's attitudes toward the public schools. Phi Delta Kappan, 66(1), 23-38.

Gallup, A. M. (1985). The 17th annual Gallup poll of the public's attitudes toward the public schools. Phi Delta Kappan, 67(1), 35-47.

Gallup, A. M. (1986). The 18th annual Gallup poll of the public's attitudes toward the public schools. Phi Delta Kappan, 68(1), 43-59.

Grading teachers: Tough enough? (1986, July 13). Washington Post, p. B6

McCormick, E. J. (1976). Job and task analysis. In M. D. Dunnette (Ed.), Handbook of industrial and organizational psychology (pp. 651-696). Chicago, Il: Rand McNally College Publishing Company.

Murnane, R. J. & Cohen, D. K. (1986). Merit pay and the evaluation problem: Why most merit pay plans fail and few survive. Harvard Educational Review, 56(1), 1-17.

Murray S. L. (1986). Policy options for teacher testing. Portland, OR: Northwest Regional Eductional Laboratory. Paper prepared for the Office of Educational Research and Improvement, U.S. Department of Education under contract number 400-86-0006.

Pecheone, Raymond L., Tomala, Gail & Forigione, Pascal D. Jr. (1986). Building a competency test for prospective teachers. In W. P. Gorth & M. L. Chernoff (Ed.), Testing for teacher certification (pp. 99-113).

Popham, W. James (1985, Fall). Recertification tests for teachers. Educational Measurement: Issues and Practice, 4(3), 23-25.

Rodman, B. (1986, May 14). In Texas, nearly 97 percent pass educators' test. Education Week, pp. 1, 40.

Rosenfeld, M., Thornton, R. F. & Skurnik, L. S. (1986). Analysis of the professional functions of teachers: Relationships between job functions ad the NTE core battery (Research Report No. 86-8). Princeton, NJ: Educational Testing Service.

Rothman, R. (1986, September 24). "Startling" data upset certification program for science teachers. Education Week, pp. 1, 15.

Schalock, H. D. (1979). Research on teacher selection. In D. C. Berliner, Annual Review of Research in Education, Vol. VII.

Schulman, L. S. (1986). Those who understand: Knowledge growth in teaching. Educational Researcher, 4515 (2), 4-14.

Shanker, Albert (1985, Fall). A national teacher examination. Educational Measurement: Issues and Practice, 4(3), 28-31.

Shimberg,B. (1981). Testing for licensure and certification. American Psychologist, 36, 1138-1146.

The Holmes Group, Inc. (1986). Tomorrows teachers: A report of the Holmes group. East Lansing, MI: Author.

U.S. Department of Health, Education, and Welfare, Public Health Services (1977). Credentialing health manpower (DHEW Publication No. (05) 77-50057). Washington, D.C.: Author.

Vorwerk, Katherine E., & Gorth, William P. (1986). Common themes in teacher certification testing program development and implementation. In W. P. Gorth & M. L. Chernoff (Ed.), Testing for teacher certification (pp. 35-43).

Northwest Regional Educational Laboratory

Dr. Robert R. Rath, Executive Director
Dr. Ethel Simon-McWilliams, Associate Director

The Northwest Regional Educational Laboratory (NWREL) is an independent, nonprofit research and development institution established in 1966 to assist education, government, community agencies, business and labor in improving quality and equality in educational programs and processes by:

- Developing and disseminating effective educational products and procedures
- Conducting research on educational products and procedures
- Providing technical assistance in educational products and procedures
- Evaluating effectiveness of educational programs and projects
- Providing training in educational planning, management, evaluation and instruction
- Serving as an information resource on effective educational programs and processes including networking among educational agencies, institutions and individuals in the region

Programs
Center for Advancement of Pacific Education
John Kofel, Director
Center for National Origin, Race, and Sex Equity
Ethel Simon-McWilliams, Director
Education and Work
Larry McClure, Director
Evaluation and Assessment
Gary Estes, Director
Literacy and Language
Stephen Reder, Director
R&D for Indian Education
Joe Coburn, Director
Rural Education
Steve Nelson, Director
School Improvement
Bob Blum, Director
Technology
Don Holznagel, Director
Western Center for Drug-Free Schools and Communities
Judith A. Johnson, Director

Program Support Services
Program and Service Coordination
Rex Hagens, Director
Institutional Development and Communications
Jerry Kirkpatrick, Director
Finance and Administrative Services
Joe Jones, Director

NWREL Headquarters
101 S.W. Main Street, Suite 500
Portland, Oregon 97204
503-275-9500
SOURCE STL058

Center for Advancement of Pacific Education
1164 Bishop Street, Suite 1409
Honolulu, Hawaii, 96813
(808) 533-1748

Alaska Office
Goldstein Building, Room 506
130 Seward Street
Juneau, Alaska 99801
(907) 586-4952

Rocky Mountain Office
1860 Lincoln Street, Suite 320
Denver, Colorado 80295
(303) 830-3675